Вестник Донского государственного технического университета. 2020. Т. 20, № 2. С. 196–200. ISSN 1992-5980 eISSN 1992-6006
Vestnik of Don State Technical University. 2020. Vol. 20, no. 2, pp. 196–200. ISSN 1992-5980 eISSN 1992-6006

ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ
INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND MANAGEMENT



УДК 004.414.23 https://doi.org/10.23947/1992-5980-2020-20-2-196-200

Искусственный интеллект в системах хранения данных



В. В. Жилин[1], О. А. Сафарьян[2]
[1] ФГКВОУ ВО «Военная академия связи имени Маршала Советского Союза С. М. Будённого», (г. Санкт-Петербург, Российская Федерация)
[2] ФГБОУ ВО «Донской государственный технический университет», (г. Ростов-на-Дону, Российская Федерация)

Введение. Рассмотрено функционирование искусственного интеллекта (ИИ) в системах хранения данных. Определено преимущество его использования при работе с данными как с экономической точки зрения, так и с точки зрения безопасности. Целью работы является внедрение искусственного интеллекта в системы хранения данных. Основные задачи: описание методов разделения данных, организации их хранения и противодействия угрозам безопасности.
Материалы и методы. Данные, которые необходимо занести на накопители, разбивается на части таким образом, чтобы их можно было восстановить, не имея одной из частей. Это необходимо для того, чтобы иметь возможность доступа и восстановления информации в случае программного или аппаратного сбоя.
Результаты исследования. Рассмотрена работа искусственного интеллекта при обнаружении угроз безопасности. Так как модель подразумевает взаимодействие пользователей с данными, то было выяснено, каким образом происходит управление доступом данных, а также приведено описание способа хранения ключей.
Обсуждение и заключения. Использование искусственного интеллекта при организации хранилища данных позволит увеличить скорость работы системы. Искусственный интеллект с встроенными алгоритмами машинного обучения позволит реагировать на ситуацию, влияющую на состояние системы. Анализ состояния накопителей позволит избежать возможного аппаратного или программного сбоя. Минимизация человеческого фактора в функционировании системы способствует улучшению её работы.

Ключевые слова: искусственный интеллект, пороговое разделение, угроза, машинное обучение, организация хранилища, динамическое изменение, ключ, шифрование, резервное копирование, атака, хеш-сумма.

Образец для цитирования*:* Жилин, В. В. Искусственный интеллект в системах хранения данных / В. В. Жилин, О. А. Сафарьян // Вестник Донского государственного технического университета. — 2020. — Т. 20, № 2. — С. 196–200. https://doi.org/10.23947/1992-5980-2020-20-2-196-200



Artificial intelligence in data storage systems

V.V. Zhilin[1], O.A. Safar'yan[2]
[1] Budenny Military Academy of Communication (St. Petersburg, Russian Federation)
[2] Don State Technical University (Rostov-on-Don, Russian Federation)

Introduction. The artificial intelligence (AI) performance in data storage systems is considered. When working with data, the advantage of its use both in economic terms and for security is determined. The work objective is the introduction of artificial intelligence in data storage systems. The key tasks involve the description of methods for data separation, organization of its storage and counteraction to security threats.
Materials and Methods. The data that should be fed into the drives is divided into parts so that it can be restored without one of the parts. This is necessary to be able to access and recover information in the event of a software or hardware failure.
Results. The AI performance under detecting security threats is considered. Since the model implies the interaction of users with data, it was found out how the data access control is carried out and the keys are stored.
Discussion and Conclusions. The use of AI in organizing a data warehouse will speed up the system. Artificial intelligence with built-in machine-learning algorithms will provide responding to a situation that affects the state of the sys-

tem. Analysis of the state of the drives will avoid a possible hardware or software failure. Minimization of the human factor in the system operation contributes to the improvement of its work.

Keywords: artificial intelligence, threshold separation, threat, machine learning, storage organization, dynamic change, key, encryption, backup, attack, hash.

For citation: V. V. Zhilin, O. A. Safar'yan . Artificial intelligence in data storage systems. Vestnik of DSTU, 2020, vol. 20, no. 2, pp. 196–200. https://doi.org/10.23947/1992-5980-2020-20-2-196-200

Введение. На современном этапе развития информационных и телекоммуникационных технологий человек встречается с различного рода информацией. Актуальным является вопрос хранения этих данных в безопасном месте. Хранением информации в зашифрованном виде занимается криптография — наука, изучающая способы сокрытия данных и обеспечения их конфиденциальности[1].

Во всех реализациях хранения данных информация содержится статично. То есть, данные будут находиться там, где они оказались первый раз. На другой диск или другой сектор информация может попасть только после удаления её с первоначального местоположения. Однако данный факт можно выделить как недостаток существующих алгоритмов.

Для повышения эффективности защиты рекомендуется проводить разделение данных [1] с целью хранения их частей. Это позволяет организовать отказоустойчивые хранилища, в которых организованы алгоритмы восстановления информации в случае какого-либо сбоя, выхода из строя одного из накопителей, а также в случае утери одной или нескольких частей данных.

Если злоумышленник имеет доступ к диску, он сможет получить необходимую информацию. При статичном хранении в случае получения доступа к нескольким таким дискам, нелегитимный пользователь может воссоздать информацию. Хранение данных в разобщенном виде повышает защищенность информации. Кроме того, в больших хранилищах, как правило, используются накопители на жестких магнитных дисках, у которых скорость работы (проведения операций чтения и записи) существенно ниже, чем у твердотельных SSD и флэш-накопителей [2]. В связи с этим приходится выбирать между используемым объёмом и скоростью работы.

Цель работы — описать функционирование искусственного интеллекта в системах хранения данных. В работе поставлена задача — описать алгоритмы работы ИИ при выполнении следующих операций:

- запись данных на накопители;
- обращение пользователей;
- анализ хранилища данных;
- в случае угроз безопасности информации.

Влияние искусственного интеллекта на качество работы системы хранения. В современных устройствах, таких как смартфоны и компьютеры, разработчики уделяют особое внимание внедрению искусственного интеллекта. В данных технологиях реализованы алгоритмы машинного обучения, что увеличивает скорость работы и сокращает время отклика на проведение часто повторяющейся информации. Следует рассмотреть, каким образом внедрение искусственного интеллекта и машинного обучения в системы хранения позволят существенно повысить качество их работы [2].

В настоящее время крупные компании и организации начинают внедрять ИИ в свои хранилища данных. Широкое внедрение технологий машинного обучения и искусственного интеллекта способствует повышению качества работы на уровне управления. Это облегчает работу администраторам сети и хранилища данных путем постоянного диагностирования причин перегрузок и снижения трафика, что позволит им заблаговременно определять потенциально уязвимые сегменты используемой модели.

Искусственный интеллект подразумевает использование интегрированных алгоритмов глубокого обучения, которые смогут прогнозировать состояние всей системы и оперативно реагировать на возможные изменения. Это позволит существенно сократить расходы по ликвидации последствий, вызванных выходом из строя оборудования. Кроме того, внедрение искусственного интеллекта в организацию отказоустойчивых хранилищ позволит обеспечить их автоматизацию [3]. Под этим подразумевается анализ состояния системы и обработка поступающих данных в динамическом режиме.

Искусственный интеллект и машинное обучение позволят минимизировать вероятность потери данных. В совокупности с избыточными массивами независимых дисков такая система увеличивает доступность и

[1] Рябко Б. Я., Фионов А. Н. Криптография в информационном мире. М., 2018. 305 с. URL: https://www.techbook.ru/book.php?id_book=1001 (дата обращения: 04.05.2020).
[2] Сравнение SSD и HDD дисков в реальных условиях использования / Хабр. URL: https://habr.com/ru/post /394135/ (дата обращения: 04.05.2020).

скорость выхода из вынужденного простоя благодаря интеллектуальному восстановлению данных, стратегии резервного копирования и переноса необходимых данных [4].

Распределение данных по входным параметрам. Для хранения данных в разделенном виде могут быть использованы методы порогового разделения. В классических алгоритмах входные параметры — статичные величины, что является существенным недостатком. Получение доступа к одним данным путём подбора входных параметров ставит под угрозу все остальные данные в системе.

Таким образом, наибольшее предпочтение в рамках информационной безопасности отдается тому алгоритму, который использует различные входные параметры для порогового разделения. Эти параметры могут генерироваться случайным образом по определенному алгоритму или зависеть непосредственно от входных данных, которые будут соответствующим образом проанализированы для подбора наиболее предпочтительных параметров. Генерацией таких параметров будет заниматься непосредственно искусственный интеллект. При этом он должен учитывать количество пользователей, которым доступна эта информация. Такие данные как параметры порогового разделения и местонахождение первой доли хранятся в базе. Эта база данных находится в зашифрованном виде на ключах конкретных пользователей, что соответствует использованию ассиметричного алгоритма шифрования.

Параметры предлагаемого алгоритма по обеспечению надежности данных. Искусственный интеллект при распределении данных по накопителям использует следующие параметры: скорость работы накопителей, их доступность, свободный объем и показатель надёжности. При этом ИИ для вычисления использует параметры ценности данных, размер и частоту обращений к ним.

Алгоритм распределения долей данных по накопителям основан на вычислении коэффициентов накопителей. Для определения скорости работы накопителя S необходимо найти среднее арифметическое скорости записи $s_{зап}$ и скорости чтения $s_{чт}$:

$$S = \frac{s_{зап} + s_{чт}}{2}.$$

Доступность накопителя A (availability) имеет 3 уровня: 0 — недоступен, 1 — частая ситуация недоступности накопителя, 2 — редкая ситуация недоступности накопителя, 3 — постоянно доступен. Показательный коэффициент накопителя может быть выражен по формуле:

$$K_{н} = S \cdot A \cdot V \cdot R,$$

где V — объём диска; R (reliability) — надёжность, она ранжируется от 1-го до 10-и баллов; A — количество обращений к файлу за определённый промежуток времени.

Коэффициент значимости файла $K_{ф}$ зависит от уровня ценности (1 — низкая, 2 — средняя, 3 — высокая). На основании известных атрибутов этот параметр можно вычислить по формуле:

$$K_{ф} = S \cdot V \cdot A.$$

Рассмотрим способ выбора накопителя хранения остальных файлов. Для определения приоритета накопителя используем номер файла n_1 в отсортированной таблице по убыванию параметра $K_{ф}$. Тогда номер диска N_1, участвующего в выборке для хранения первой доли данных будет найден по формуле:

$$\boldsymbol{N_1 = round\left(\frac{d \cdot n_1}{f}\right) mod\; n_2},$$

где d — количество дисков, используемых для хранения; f — количество файлов в системе; n_2 — количество накопителей; $round$ — округление к ближайшему целому.

Таким же образом определяется диск N_2, который возможно будет выбран в качестве накопителя для хранения первой доли:

$$K_{first} = K_{ф} \cdot K_{d.max},$$

где $K_{ф}$ — коэффициент разбиваемого файла; $K_{d.max}$ — максимальный коэффициент накопителя.

Далее в качестве N_2 выбирается тот накопитель, у которого разница коэффициентов K_{first} и $K_{н}$ минимальна. Для хранения первой доли случайным образом выбирается один из найденных дисков N_1 и N_2.

Преимущества рассмотренного алгоритма распределения данных по накопителям. Классически диски, объединенные в один массив данных (*RAID*-массив), работают по статичному алгоритму. То есть при обнаружении первой доли данных можно провести процедуру поиска второй доли. Каждый раз, при обнаружении очередной доли вероятность определить конкретный алгоритм распределения повышается. В предложенном алгоритме распределение происходит на основе информационных данных и накопителей. Другими словами, задача нахождения всех долей данных является *NP*-полной, то есть не решаемой за полиномиальное время. Из этого следует, что использование данного алгоритма повышает надежность системы хранения информационных данных.

Однако, пользователи не создают ключ, он генерируется в автоматическом режиме и хранится на устройствах самих пользователей. Таким образом, при попытке расшифровать данные должен использоваться конкретный ключ, иначе данная операция завершится неудачно. Местонахождение ключей изменяется динамически с учетом времени. Такие изменения могут происходить в определенные часы, либо через какой-либо промежуток времени [5]. Искусственный интеллект отвечает за местоположение ключа с учетом доступности всех устройств, подключенных к системе. Это позволяет усложнить работу злоумышленнику, чьей целью является доступ к информации, хранящейся на накопителях [6].

Так как информацией, находящейся в хранилище, оперируют пользователи, следует определять, кому из них разрешено проводить операции с данными, другими словами, как разграничивать доступ к данным.

Сравнение доступа происходит по таблице хеш-сумм [1]. Если у пользователя есть доступ к данным, то они восстанавливаются. В случае отсутствия записи в базе данных о предоставлении прав доступа, субъекту, запросившему доступ к объекту, в нём будет отказано. Кроме того, попытка получения доступа будет зафиксирована в лог-файле, а субъект, которому принадлежит хранимая в системе информация, будет проинформирован о попытке получения доступа к его данным нелегитимным пользователем или субъектом. Искусственный интеллект в данном случае анализирует действия каждого пользователя с целью принятия того или иного решения в случае возникновения определенной ситуации. Например, он определяет, является ли запрос ошибочным или имеет какую-либо цель [2].

На рис. 1 представлен пример предоставления доступа на восстановление данных и отказ в доступе с занесением события в лог-файл и уведомлением пользователя.

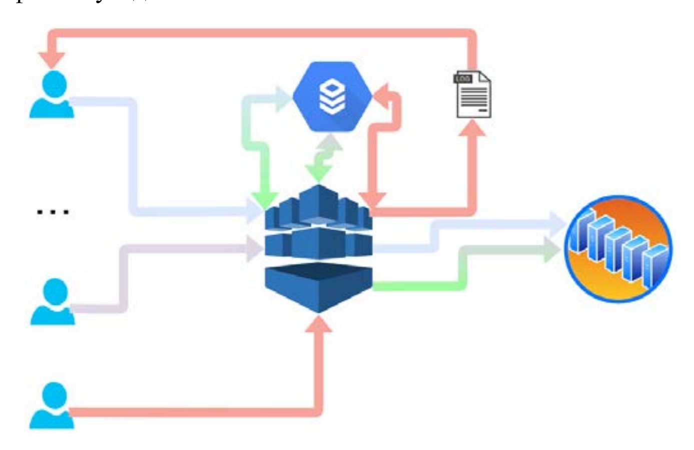


Рис. 1. Предоставление и отказ доступа на восстановление данных

Рассмотрим функцию управления доступом. Так как в системе основными субъектами являются пользователи, то вопрос предоставления доступа к данным является особо актуальным. Очень важен полный контроль над операциями пользователей. Решение о предоставлении доступа к данным принимает распределяющее устройство с поддержкой искусственного интеллекта. Использование ИИ в данном случае позволит увеличить скорость работы пользователя над его данными. Это достигается путём трансформации наиболее часто используемой пользователем информации, что позволяет избежать ожидания времени отклика на отправленный запрос о восстановлении данных.

Рассмотрим динамическое изменение хранилища. В определенных случаях искусственный интеллект управляет переносом данных и отвечает за резервное копирование и восстановление данных, в т. ч. с учетом времени. При этом данные мигрируют с одного диска на другой, а информация в базе данных также изменяется. Это позволяет минимизировать вероятность прогнозирования состояния системы в конкретный момент времени. Так как все ключи генерируются на основе предыдущих последовательностей с использованием хеш-сумм, задача нахождения ключа, предназначенного для расшифровывания базы данных является *NP*-полной задачей [7]. Искусственный интеллект в данном случае анализирует состояние накопителей, а также системы в целом. При обнаружении неисправности происходит перенос данных с тех дисков, на которых замечены сбои. Это позволит минимизировать вероятность потери информации.

При обнаружении угрозы искусственный интеллект анализирует атаку с целью обнаружения её цели. Если конечной целью является объект хранилища, то данные с него переносятся на другой накопитель. Таким

[1] Легкий способ подделки контрольной суммы с помощью коллизий / xakep.ru. URL: https://xakep.ru /2012/11/22/light-fake-checksum/ (дата обращения: 04.05.2020).

образом, в случае успешной атаки, ИИ будет предпринимать эффективные действия для сокрытия информации [8].

Поскольку с данными работают пользователи, то важную роль играет человеческий фактор. Поэтому система устроена таким образом, чтобы с данными могли работать только те пользователи, у которых имеется доступ. Здесь есть аналогия с мандатной моделью управления доступом. Исключением является то, что раздачей доступа управляет пользователь, создавший информацию в системе и являющийся её владельцем.

Заключение. Таким образом, использование искусственного интеллекта при организации хранилища данных повышает скорость работы системы. Ограничение доступа пользователей в алгоритме работы системы улучшает информационную безопасность. Искусственный интеллект с встроенными алгоритмами машинного обучения позволит оперативно реагировать на любую ситуацию, влияющую на состояние системы. Анализ состояния накопителей позволит избежать возможного аппаратного или программного сбоя. Минимизация человеческого фактора в работе системы способствует улучшению её функционирования и более глубокому анализу пользовательских запросов. Кроме того, сбор информации о возможных атаках позволит поддерживать на должном уровне безопасность системы.

Библиографический список

1. Могилевская, Н. С. Пороговое разделение файлов на основе битовых масок: идея и возможное применение / Н. С. Могилевская, Р. В. Кульбикаян, Л. А. Журавлёв // Вестник Донского государственного технического университета : [сайт]. — 2011. — Т. 11, № 10. — С. 1749–1755. — URL: https://vestnik.donstu.ru/jour/article/ view/912/907 (дата обращения: 04.04.2020).
2. Николенко, С. И. Глубокое обучение. Погружение в мир нейронных сетей / С. И. Николенко, А. А. Кадурин, Е. В. Архангельская. — Санкт-Петербург : Питер, 2018. — 481 с.
3. Dubrova, E. Fault-Tolerant Design / Springer, 2013. — 185 p.
4. Флах, П. Машинное обучение / П. Флах. — Москва : ДМК Пресс, 2015. — 400 с.
5. Трехмерная модель безопасности компьютерных систем / В. В. Жилин, И. И. Дроздова, Л. В. Черкесова, О. А. Сафарьян // Молодой исследователь Дона : [сайт]. — 2018. — № 5. — С. 30–37. — URL: http://mid-journal. ru/upload/iblock/f81/6_620_ZHilin_30_37.pdf (дата обращения: 04.05.2020).
6. Parloff, R. Why Deep Learning Is Suddenly Changing Your Life / R. Parloff // Fortune. — 2016. (Retrieved 13 April, 2018.).
7. Алгоритмы: построение и анализ / Кормен Томас Х., Лейзерсон Чарльз И., Ривест Рональд Л. — Москва : Вильямс, 2006. — 1296 с.
8. Hutson, M. Missing data hinder replication of artificial intelligence studies / Matthew Hutson // Science. — 15 February, 2018. doi:10.1126/science.aat3298.

Сдана в редакцию 03.02.2020
Запланирована в номер 17.03.2020

Об авторах:

Жилин Виктор Владимирович, студент кафедры «Кибербезопасность информационных систем», ФГКВОУ ВО «Военная академия связи имени Маршала Советского Союза С. М. Будённого», (194064, РФ, г. Санкт-Петербург, пр. Тихорецкий, 3), ORCID: https://orcid.org/0000-0001-6277-3795, zhilin95@inbox.ru

Сафарьян Ольга Александровна, доцент кафедры «Кибербезопасность информационных систем», ФГБОУ ВО «Донской государственный технический университет», (344000, РФ, г. Ростов-на-Дону, пл. Гагарина, 1), кандидат технических наук, доцент, ScopusID 57210832767, ORCID: https://orcid.org/0000-0002-7508-913X, safari_2006@mail.ru

Заявленный вклад соавторов

В. В. Жилин — сбор и анализ литературных данных, определение методологии исследования, постановка задачи. О. А. Сафарьян — научное руководство, формулирование основной концепции исследования и структуры статьи.

Все авторы прочитали и одобрили окончательный вариант рукописи.